DOCUMENT RESUME

ED 343 776 SE 052 278

AUTHOR Silver, Edward A.; Lane, Suzanne
TITLE Assessment in the Context of Mathematics Instruction Reform: The Design of Assessment in the QUASAR Project.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
PUB DATE Apr 91
NOTE 12p.; Paper presented at the Meeting of the International Commission on Mathematical Instruction on Assessment in Mathematics Education and Its Effects (Calogne, Spain, April 1991).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Economically Disadvantaged; Educational Strategies; Elementary School Mathematics; *Evaluation; Instructional Development; *Instructional Improvement; Intermediate Grades; Junior High Schools; Mathematics Education; *Mathematics Instruction; Mathematics Tests; *Middle Schools; Poverty; Program Descriptions; *Program Development; Teaching Methods
IDENTIFIERS *Mathematical Power; Mathematics Education Research; Middle School Students; QUASAR Project (Mathematics Education)

ABSTRACT

Recent reports on mathematics education reform have focused the attention of educational practitioners and policymakers on new goals for mathematics education and new descriptions of mathematical proficiency. QUASAR is a national project (Quantitative Understanding: Amplifying Student Achievement and Reasoning) designed to improve the mathematics instructional program for students attending middle schools, grades 6 through 8, in economically disadvantaged communities. QUASAR is a complex research study of educational change and improvement, in which a major effort will be made to study carefully different approaches to unblocking the path to mathematical power for poor students. Parallel goals for the study are: to ascertain conditions that appear conducive to mathematical success; to derive pedagogical principles for effective mathematics instruction for middle school students; to describe effective instructional programs that are adaptable to other schools; and to devise new assessment tools to measure growth in higher order thinking, reasoning, and communication as they relate to school mathematics. Included in this report are: (1) an introduction that describes the purpose, the rationale, and the goals of this project; (2) a discussion of the educational considerations and mathematical conceptualizations underlying the proposed methods of assessment for mathematical proficiency; (3) a discussion of construct-irrelevant test variance as a data-gathering consideration for the assessment of mathematical proficiency; (4) a discussion of the development of specifications for the assessment tasks in terms of focus and components; (5) a discussion of the specifications encompassing the scoring rubrics within the assessment procedures; and (6) a list of sample tasks and administrative information. (15 references) (Author/JJK)

ED343776

Assessment in the Context of Mathematics Instruction Reform: The Design of Assessment in the QUASAR Project

Edward A. Silver

Suzanne Lane

Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Aug

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: QUASAR Assessment

Paper presented at the International Commission on Mathematical Instruction conference on Assessment in Mathematics Education and Its Effects, Cologne, Spain, April 1991

SE 052 278

BEST COPY AVAILABLE

Assessment in the Context of Mathematics Instruction Reform: The Design of Assessment in the QUASAR Project

Edward A. Silver
Suzanne Lane

Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260

Mathematics education reform is currently a topic of great interest in the United States. Reports by the National Academy of Sciences (National Research Council, 1989), the American Association for the Advancement of Science (1989) and the National Council of Teachers of Mathematics (1989) have focused the attention of educational practitioners and policy makers on new goals for mathematics education and new descriptions of mathematical proficiency. Terms like reasoning, communication, problem solving, conceptual understanding, and mathematical power are used frequently to describe an expanded view of mathematical proficiency that goes beyond memorization and mere competence in the basic skills of rational number computation. The reform discussion has thus led naturally to considerations of how to assess students' attainments with respect to this new vision of mathematical proficiency and how to assess improvements that may result from curricular and instructional reforms that might be undertaken. This paper focuses on the efforts of one project to deal with the interface between assessment and instructional reform.

QUASAR (Quantitative Understanding: Amplifying Student Achievement and Reasoning) is a national project designed to improve the mathematics instructional program for students attending middle schools (grades 6-8) in economically disadvantaged communities (Silver, 1989). Currently operating at 6 school sites dispersed across the United States (Silver, Smith, Lane, Salmon-Cox, & Stein, 1990), QUASAR is a practical school demonstration project which posits that students in these communities can and will learn a broader range of mathematical content, acquire a deeper and more meaningful understanding of mathematical ideas, and demonstrate an ability to reason and solve appropriately complex problems. When implemented, such instructional programs will stand in stark contrast to those characterized by what might be called "assembly line" mathematics instruction -- a program of repetitive drill and practice on basic computation which has characterized middle school mathematics education for many American students and which has relegated disproportionate numbers of poor students to the remedial track,

thereby blocking their access to most socially acceptable paths to status and success. QUASAR is also a complex research study of educational change and improvement, in which a major effort will be made to study carefully different approaches to accomplishing this general goal; to ascertain conditions that appear to be conducive to success; to derive instructional principles for effective mathematics instruction for middle school students; to describe effective instructional programs in ways that will allow their adaptation to other schools, and to devise new assessment tools to measure growth in high-level thinking, reasoning and communication as they relate to mathematics.

Given the goals and aspirations of the QUASAR project, it is imperative that appropriate measures be developed to monitor and evaluate program impact. One important set of indicators are those that pertain to growth in student knowledge and proficiency over time. Development of the assessments for the QUASAR project has utilized an approach advocated by the National Council of Teachers of Mathematics Curriculum and Evaluation Standards for School Mathematics (1989). That report argued for improving the alignment of testing with curriculum goals, advocated the use of multiple sources of assessment information, and suggested that more attention be given both to appropriate methods of assessment and to the proper use of assessment information. With respect to the methods of assessment, the report asserted that an authentic assessment of mathematical proficiency would need to address such areas as problem solving, communication, reasoning, and disposition, as well as concepts and procedures.

The QUASAR project will employ a variety of measures in assessing student growth, including paper-and-pencil cognitive assessment tasks administered to individual students in a large group setting; tasks administered to students in small groups, and on which they are expected to work collaboratively; individually administered performance assessments, which may involve the use of manipulative materials and computational tools; tasks designed to provide information on metacognitive processes used in problem solving; and non-cognitive assessments aimed at important attitudes, beliefs, and dispositions. Teachers at the project sites are also asked to supply information available from their own classroom sources (e.g., tests, homework, projects) to supplement the store of information about both the program and individual students.

In the development of assessments, the project has attempted to keep a balanced perspective regarding psychometric constraints and educational needs. This has been possible because the coordinator of assessment development (S. Lane) is a psychometrician by training and the project director (E. Silver) is a mathematics educator. We believe that this balanced perspective is essential for significant progress to be made in establishing alternative assessments as possible replacements for or supplements to the current system

of standardized, multiple-choice testing that has become entrenched in the United States. This paper presents an overview of the design principles for the development of the paper-and-pencil mathematics assessment instrument that is administered to individual students in a large group setting.

The QUASAR assessments are designed to provide programmatic rather than individual student information. In other words, we are not attempting to provide valid, reliable indicators for the purpose of evaluating individual students; rather, we have designed a system that will collect data from individual students but will provide evaluative information only at the program level. Therefore, a relatively large number of assessment tasks (currently about 36) is administered at each project site, but each student completes only a small number of the tasks (about 9) on each administration occasion. Because of our focus on program evaluation, use of this approach allows us to avoid the difficulty of sampling only a small range of tasks. Over time, it is planned to release some assessment tasks and add new ones. The public release of tasks and scoring rubrics should allow for a clearer understanding of the nature of mathematical proficiencies being assessed and the judgment criteria that are applied in the evaluation of responses. The addition of new tasks each year will allow the QUASAR assessment instrument to expand to include not only tasks that reflect important general instructional emphases and topics but also some tasks that have been tailored to reflect the unique features of instructional programs that vary across sites; these latter tasks could be developed in close cooperation with the teachers and resource partners at each project site.

Given the goals of the QUASAR project regarding instructional program emphases on breadth of content, tasks have been developed to assess students' knowledge across a wide range of content areas -- going well beyond whole numbers and arithmetic. Also, given the project's goals related to high-level thinking and deep conceptual understanding, the assessment tasks focus on mathematical reasoning, problem solving, and modeling, and on students' understanding of the features that characterize mathematical concepts and their interrelationships. Due to space limitations, the description of QUASAR assessment in this paper will be quite brief in some places. Further details regarding the design principles and conceptual framework for the assessment can be found in Lane (1991).

QUASAR's Assessment of Mathematical Proficiency: Some Educational Considerations

The parameters that characterize QUASAR's vision of mathematical ability and mathematical power have been described to a large extent in the Curriculum and Evaluation Standards for School Mathematics (National Council of Teachers of Mathematics, 1989), which suggest the importance of understanding concepts and procedures, becoming a

mathematical problem solver, learning to reason mathematically, making connections among mathematical topics and between mathematics and the world outside the mathematics classroom, and learning to communicate mathematical ideas. The vision is also consistent with that of the Mathematical Sciences Education Board (National Research Council, 1990) which argued that mathematical power involved the development of the abilities to understand mathematical concepts, principles and procedures, to discern mathematical relations, to reason mathematically, and to apply mathematical concepts, principles, and procedures to solve a variety of nonroutine problems.

In this view, mathematics is conceptualized as involving problems that are complex, yield multiple solutions, require judgment and interpretation, require finding structure, and require finding a path for a solution that is not immediately visible. Furthermore, success in mathematical problem solving is viewed as being related to and at least partially dependent on students' beliefs about the nature of mathematics and problem solving, attitudes towards and interest in mathematics, and the socio-cultural context (Lester & Kroll, 1990; Silver, 1985). Specifications for the QUASAR assessment tasks were based upon these conceptualizations of mathematical proficiency.

QUASAR's Assessment of Mathematical Proficiency: Some Measurement Considerations

An assessment instrument is an imperfect measure of a construct because it either underrepresents the construct domain (i.e., the assessment instrument is too narrow) or in addition to measuring the construct domain it also measures something that is irrelevant to the construct (i.e., irrelevant excess reliable variance), or some combination of the two (Messick, 1989). To ensure that the construct domain is fully represented, QUASAR's assessment of mathematical proficiency is sensitive to many facets, including mathematical reasoning, mathematical communication, knowledge and use of strategies and representations, and knowledge and use of mathematical concepts, principles, and procedures. Moreover, the assessment attends to the fact that these facets interact with various mathematical content areas such as number sense, geometry, and statistics.

Two kinds of construct-irrelevant test variance are proposed by Messick (1989): construct-irrelevant easiness and construct-irrelevant difficulty. Construct irrelevant easiness refers to the potential of clues or flaws in task format which may allow some students to respond correctly in ways that are irrelevant to the construct domain being measured, and which may lead to scores that are invalidly high. Construct-irrelevant difficulty refers to the possibility that the assessment instrument is, for irrelevant reasons, more difficult for some groups of students. In QUASAR's assessments of students' abilities to think and reason mathematically, we were sensitive to several potential irrelevant

constructs that could adversely affect some groups of students, such as differences in reading comprehension ability, writing ability, or familiarity with task contexts. Therefore, the degree of reading and writing required of the student by the task was considered in developing open-ended assessment tasks and scoring rubrics, as was the likely familiarity of the task contexts to students of differing cultural and ethnic backgrounds. Not only were these two sources of invalidity considered in the process of constructing the assessment tasks and corresponding scoring rubrics but they will also be considered when interpreting student performances.

Another measurement issue relates to the reliance on a single measure of a complex construct. To triangulate observations of a complex construct, multiple measures are needed. To measure program outcomes and growth in the QUASAR project, the core assessment instrument incorporates a number of task formats (e.g., requiring a student to justify a selected answer vs. showing the solution process used to arrive at an answer) and process constraints (e.g., producing a numerical answer vs. drawing a diagram). Moreover, as Baker (1990) has noted, any measurement procedure must be understood in the light of other available information and the intended uses of the scores. Therefore, information will also be obtained about classroom processes, students' class assignments and assessments, teachers' knowledge and beliefs about mathematics, and students' beliefs about and disposition towards mathematics.

Specification of the Assessment Tasks

The development of QUASAR's assessment tasks and scoring rubrics involves a collaborative effort by a team consisting of mathematics educators, mathematicians, cognitive psychologists, and psychometricians. Our approach is related to but somewhat different from other examples of alternative assessment frameworks (e.g., Nitko & Lane, 1990; Pandey, 1990; Romberg, Zarinnia, & Collis, 1990). The assessment tasks are specified in terms of four components: cognitive processes, mathematical content, mode of representation, and task context. With a particular focus on mathematical problem solving and mathematical reasoning, the cognitive processes that were specified for task development included the following: understanding and representing problems, discerning mathematical relations, organizing information, using and discovering strategies and heuristics, using and discovering procedures, formulating conjectures, evaluating the reasonableness of answers, generalizing results, and justifying answers or procedures. The content categories included the following: number and operations (involving decimals, fractions, ratios, and proportions); estimation (both computational and measurement); patterns (both numerical and geometric/spatial patterns); algebra (especially tasks related to

the transition from arithmetic to algebra); geometry and measurement; and data analysis (including probability and statistics). The types of representations used in task development and expected of students in developing the scoring rubrics include written, pictorial, graphic, tabular, and arithmetic representations. With respect to task context, an attempt was made to embed as many tasks as possible within an appropriate context if it could be done without requiring an excessive amount of reading on the part of the students.

Specification of Scoring Rubrics

A focused holistic scoring method is being used to score students' responses to each task. A generalized scoring rubric was designed to incorporate three interrelated components related to the task development specifications described above: mathematical conceptual and procedural knowledge, strategic knowledge, and communication. With respect to mathematical knowledge, attention is paid to the extent to which students demonstrate their knowledge of mathematical concepts, principles and procedures, such as understanding relationships among problem elements; using appropriate mathematical terminology or notation; recognizing when a procedure is appropriate; executing procedures; verifying results of procedures; and generating or extending familiar procedures. In the area of strategic knowledge, students are expected to use models, diagrams, and symbols to represent and integrate concepts in addition to being systematic in their application of strategies. The area of communication relates to students' ability to communicate their mathematical ideas in writing, symbolically, or visually; to use mathematical vocabulary, notation, and structure to represent ideas; and to describe relationships and model situations. Some tasks require the justification of answers through the use of appropriate modes of communication (e.g., written, pictorial, graphical, or algebraic methods) for expressing the integration of mathematical ideas, conjectures, and arguments; other tasks require the description of strategies or patterns.

The scoring rubrics developed by the California Assessment Program (California State Department of Education, 1989) provided a basis for the development of QUASAR's generalized rubric. In developing the generalized scoring rubric, criteria representing the three interrelated components were specified for each of five score levels (0-4). Based on the specified criteria at each score level, a specific rubric was developed for each task. The emphasis on each component for a specific rubric was dependent upon the demands of the task. In addition to scoring the student responses using the scoring rubric developed for each task, the student responses will be evaluated using other more analytic procedures. These latter analyses should provide more detailed information regarding the types of representations and strategies students use, the nature of errors or misconceptions in

students' work, and the nature of the mathematical knowledge and cognitive processes underlying successful performance.

Sample Tasks and Administration Information

For the 1990-91 school year, a set of thirty-six assessment tasks was developed for use with sixth-grade students. The thirty-six tasks were divided into four sets of nine different tasks, which were randomly distributed to students in each classroom. Students received a different set in each of the Fall and Spring administrations. Two examples of assessment tasks similar to those used in the QUASAR project are provided in Figure 1.

For the first task, it is expected that a student would draw a 9-by-9 square on the grid provided and shade the square in. Also it is expected that a student would describe the pattern by saying "It is a pattern of squares with odd sides - 1, 3, 5, 7, 9, 11, and so on;" or "In the pattern you add 2 rows and 2 columns to each square to get the next square;" or some other similar description. In the next task, we would expect that a student's response would show evidence of a clear reasoning process. For example, a student might answer "no" and provide an explanation, such as "Yvonne takes the bus eight times in the week, and this would cost $8.00. Since the bus pass costs $9.00, she should not buy the pass." It is possible, however, that a student might answer "yes" and provide a logical reason, such as "Yvonne should buy the bus pass because she rides the bus eight times for work and this costs $8.00. If she rides the bus on weekends (to go shopping, etc.), it would cost $2.00 or more, and that would be more than $9.00 altogether, so she can save money with the bus pass." As this example suggests, tasks presented in this open-ended format may allow for more than one possible correct answer.

After student responses have been obtained, the papers are scored by teams of classroom teachers who are trained as raters. The raters use the scoring rubric for each task in order to assign a score between 0 and 4 to each student's response. In addition to these holistic judgements, student responses will be subjected to further examination and analysis in order to probe for systematic error patterns, cognitive process information, data regarding strategy usage, and other important insights related to the mathematical knowledge and performance of the students.

As noted earlier, QUASAR intends to use a wide range of assessment procedures. In addition to open-ended tasks similar to those shown in Figure 1, QUASAR will also utilize some performance assessments involving use of manipulative materials or computational tools, such as calculators. Performance assessments have been developed and will be implemented on a pilot basis during the 1990-91 school year. Tasks assessing students working in small groups are also planned for the near future.

References

American Association for the Advancement of Science (1989). Project 2061: Science for all Americans. Washington, DC: Author.

Baker, E. L. (1990). Developing comprehensive assessments of higher order thinking. In. G. Kulm (Ed.), Assessing higher order thinking in mathematics (pp. 7-20). Washington, DC: American Association for the Advancement of Science.

California State Department of Education (1989). A question of thinking: A first look at students' performance on open-ended questions in mathematics. Sacramento, CA: Author.

Lane, S. (1991, April). The conceptual framework for the development of a mathematics assessment instrument for QUASAR. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Lester, F.K., Jr. & Kroll, D.L. (1990). Assessing student growth in mathematical problem solving. In. G. Kulm (Ed.), Assessing higher order thinking in mathematics (53-70). Washington, DC: American Association for the Advancement of Science

Mathematical Sciences Education Board (1990). Reshaping school mathematics: A philosophy and framework for curriculum. Washington, DC: National Academy of Sciences.

Messick, S. (1989). Test validity. In R.L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 13-104). New York: American Council on Education.

National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: NCTM.

National Research Council (1989). Everybody counts. Washington, DC: National Academy of Sciences.

Nitko, A.J., & Lane, S. (1990, August). Solving problems is not enough: Assessing and diagnosing the ways in which students organize. Paper presented at the Third International Conference on Teaching Statistics, Dunedin, New Zealand.

Pandey, T. (1990). Power items and the alignment of curriculum and assessment. In. G. Kulm (Ed.), Assessing higher order thinking in mathematics (pp. 39-52). Washington, DC: American Association for the Advancement of Science.

Romberg, T.A., Zarinnia, E. A., Collis, K.F. (1990). A new world view of assessment in mathematics. In. G. Kulm (Ed.), Assessing higher order thinking in mathematics (pp. 21-38). Washington, DC: American Association for the Advancement of Science.

Silver, E. A. (1985). Research on teaching mathematical problem solving: Some underrepresented themes and needed directions. In E. A. Silver (Ed.), Teaching and learning mathematical problem solving: Multiple research perspectives (pp. 247-266). Hillsdale, NJ: Lawrence Erlbaum Associates.

Silver, E. A. (1989). QUASAR. The Ford Foundation Letter, 20(3), 1-3.

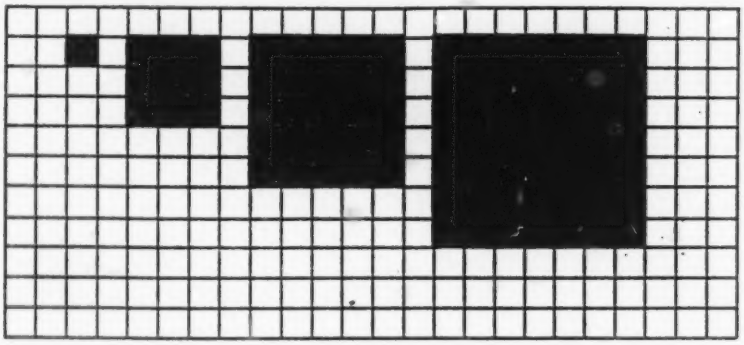
Silver, E.A., Smith, M.S., Lane, S., Salmon-Cox, L., Stein, M.K. (1990, Fall). QUASAR (Quantitative Understanding: Amplifying Student Achievement and Understanding) project summary. Learning Research and Development Center, University of Pittsburgh.

Preparation of this paper was supported by a grant from the Ford Foundation (grant number 890-0572) for the QUASAR project. Any opinions expressed herein are those of the authors and do not necessarily reflect the views of the Ford Foundation.

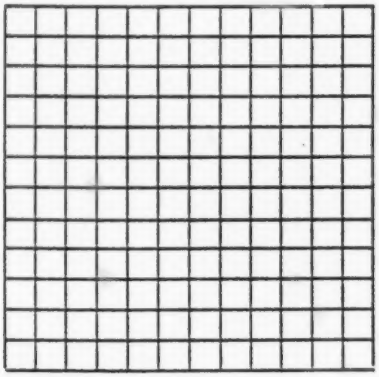
Figure 1

Sample Assessment Tasks

Task 1 - Mathematical Content: Pattern recognition
Look at the following pattern of figures:



A. Draw the 5th figure:



B. Describe the pattern.

Task 2 - Mathematical Content: Numbers and Operations
The table below shows the cost for different bus fares.

BUSY BUS COMPANY
FARES

One Way	$ 1.00
Weekly Pass	$ 9.00

Yvonne is trying to decide whether she should buy a weekly bus pass. On Monday, Wednesday and Friday she rides the bus to and from work. On Tuesday and Thursday she rides the bus to work, but gets a ride home with her friends.

Should Yvonne buy a weekly bus pass? _____

Explain your answer.

BEST COPY AVAILABLE